Bracketing Numbers of Convex Functions on Polytopes

Charles R. Doss *
School of Statistics, University of Minnesota

Abstract

We study bracketing numbers for spaces of bounded convex functions in the L_p norms. We impose no Lipschitz constraint. Previous results gave bounds when the domain of the functions is a hyperrectangle. We extend these results to the case wherein the domain is a polytope. Bracketing numbers are crucial quantities for understanding asymptotic behavior for many statistical nonparametric estimators. Our results are of interest in particular in many multidimensional estimation problems based on convexity shape constraints.

Keywords: bracketing entropy, Kolmogorov metric entropy, convex functions, convex polytope, covering numbers, nonparametric estimation, convergence rates

Mathematics Subject Classification (2010): Primary: 52A41, 41A46; Secondary: 52A27, 52B11, 52C17 62G20

1 Introduction and Motivation

To quantify the size of an infinite dimensional set, the pioneering work of Kolmogorov and Tihomirov (1961) studied the metric covering number of the set and its logarithm, the metric entropy. Metric entropy quantifies the amount of information it takes to recover any element of a set with a given accuracy ϵ. This quantity is important in many areas of statistics and information theory; in particular, the asymptotic behavior of empirical processes and thus of many statistical estimators is fundamentally tied to the entropy of the class under consideration (Dudley, 1978).

In this paper, we are interested not in the metric entropy but the related bracketing entropy for a class of functions. Let $\mathcal{F}$ be a set of functions and let d be a metric on $\mathcal{F}$. We call a pair of functions $[l, u]$ a bracket if $l \le u$ pointwise. For $\epsilon > 0$, the ϵ-bracketing number of $\mathcal{F}$, denoted $N_{[]}(\epsilon, \mathcal{F}, d)$, is the smallest N such that there exist brackets $[l_i, u_i]$, $i = 1, \ldots, N$, such that for all $f \in \mathcal{F}$, there exists i with $l_i(x) \le f(x) \le u_i(x)$ for all x. Like metric entropies, bracketing entropies are fundamentally tied to rates of convergence of certain estimators (see e.g., Birgé and Massart (1993), van der Vaart and Wellner (1996), van de Geer (2000)). In this paper, we study the bracketing entropy of classes of convex functions. Our interest is motivated by the study of nonparametric estimation of functions satisfying

*Charles R. Doss, 224 Church St. SE, Minneapolis, MN 55455. Email: cdoss@stat.umn.edu

convexity restrictions, such as the least-squares estimator of a convex or concave regression function on $\mathbb{R}^d$ (e.g., Seijo and Sen (2011), Guntuboyina and Sen (2015)), possibly in the high dimensional setting (Xu et al., 2014), or estimators of a log-concave or s-concave density (e.g., Seregin and Wellner (2010), Koenker and Mizera (2010), Kim and Samworth (2014), Doss and Wellner (2015a,b), among others). Bracketing entropy bounds are directly relevant for studying asymptotic behavior of estimators in these contexts.

Let $D \subset \mathbb{R}^d$ be a convex set, let $v_1, \ldots, v_d \in \mathbb{R}^d$, be linearly independent vectors, let $B, \Gamma_1, \ldots, \Gamma_d$ be positive reals, and let $\mathbf{v} = (v_1, \ldots, v_d)$ and $\mathbf{\Gamma} = (\Gamma_1, \ldots, \Gamma_d)$. Then we let $\mathcal{C}(D, B, \mathbf{\Gamma}, \mathbf{v})$ be the class of convex functions φ defined on D, such that $|\varphi(x)| \le B$ for all $x \in D$, and such that $|\varphi(x + \lambda v_i) - \varphi(x)| \le \Gamma_i |\lambda|$ as long as x and $x + \lambda v_i$ are both elements of D. Let $\mathcal{C}(D, B)$ be the convex functions on D with uniform bound B and no Lipschitz constraints. For $f : D \to \mathbb{R}$, let $L_p(f) = \left(\int_D f(x)^p \, dx\right)^{1/p}$ for $1 \le p < \infty$, and let $L_\infty(f) = \sup_{x \in D} |f(x)|$. Since convex functions are Lebesgue almost everywhere two-times differentiable their entropies correspond to the entropy for twice differentiable function classes, namely $\epsilon^{-d/2}$. When D is a hyperrectangle $\prod^d [-1, 1]$, and $B = 1$, $\Gamma_i = 1$, Bronshtein (1976) and Dudley (1984), chapter 8, indeed show that $\log N(\epsilon, \mathcal{C}(D, B, \mathbf{\Gamma}), L_\infty) \lesssim \epsilon^{-d/2}$. Here, $N(\epsilon, \mathcal{F}, \rho)$ is the ϵ-covering number of $\mathcal{F}$ in the metric ρ, i.e. the smallest number of balls of ρ-radius ϵ that cover $\mathcal{F}$.

Bracketing entropies govern the suprema of corresponding empirical processes and thus govern the rates of convergence of certain statistical estimators. In many problems, including some of the statistical ones mentioned above, the classes that arise do not naturally have Lipschitz constraints, and so the class $\mathcal{C}(D, B, \mathbf{\Gamma})$ is not of immediate use. Without Lipschitz constraints, the L_∞ bracketing numbers are not bounded, but one can use the L_p metrics, $1 \le p < \infty$, instead: Dryanov (2009) and Guntuboyina and Sen (2013) found bounds when $d = 1$ and $d > 1$, respectively, for metric entropies of $\mathcal{C}(D, 1)$: they found that $\log N(\epsilon, \mathcal{C}(D, 1), L_p) \lesssim \epsilon^{-d/2}$, again with D a hyperrectangle. The $d = 1$ case (from Dryanov (2009)) was the fundamental building block in computing the rate of convergence of the univariate log-concave and s-concave MLEs in Doss and Wellner (2015a). In the corresponding statistical problems when $d > 1$, the domain of the functions under consideration is not a hyperrectangle but rather is a polytope, and thus the results of Guntuboyina and Sen (2013) are not always immediately applicable, and there is need for results on more general convex domains D. It is not immediate that previous results will apply, since D may have a complicated boundary. In this paper we are able to indeed find bracketing entropies for all polytopes D, attaining the bound

$$\log N_{[\,]}(\epsilon, \mathcal{C}(D, B), L_p) \lesssim \epsilon^{-d/2}$$

with $1 \le p < \infty$, D a polytope, and $0 < B < \infty$. Note we work with bracketing entropy rather than metric entropy. Bracketing entropies are larger than metric entropies (van der Vaart and Wellner, 1996), so bracketing entropy bounds imply metric entropy bounds of the same order. Along the way, we also generalize the

results of Bronshtein (1976) to bound the L_∞ bracketing numbers of $\mathcal{C}(D, B, \mathbf{\Gamma})$ when D is arbitrary. One of the benefits of our method is its constructive nature. We initially study only simple polytopes and in that case attempt to keep track of how constants depend on D.

This paper is organized as follows. In Section 2 we prove bounds for bracketing entropy of classes of convex functions with Lipschitz bounds, using the L_∞ metric. We use these to prove our main result for the bracketing entropy of classes of convex functions without Lipschitz bounds in the L_p metrics, $1 \le p < \infty$, which we do in Section 3. We defer some of the details of the proofs to Section 4.

2 Bracketing with Lipschitz Constraints

If we have sets $D_i \subset \mathbb{R}^d$, $i = 1, \ldots, M$, for $M \in \mathbb{N}$, and $D \subseteq \cup_{i=1}^M D_i$ then for $\epsilon_i > 0$,

$$N_{[\,]}\left(\left(\sum_{i=1}^M \epsilon_i^p\right)^{1/p}, \mathcal{C}(D,1), L_p\right) \le \prod_{i=1}^M N_{[\,]}\left(\epsilon_i, \mathcal{C}(D,1)|_{D_i}, L_p\right). \tag{1}$$

where, for a class of functions $\mathcal{F}$ and a set G, we let $\mathcal{F}|_{\mathcal{G}}$ denote the class $\{f|_G : f \in \mathcal{F}\}$ where $f|_G$ is the restriction of f to the set G. We will apply (1) to a cover of D by sets G with the property that

$$\mathcal{C}(D,1)|_G \subseteq \mathcal{C}(G,1,\Gamma)$$

for some $\Gamma < \infty$, so that we can apply bracketing results for classes of convex functions with Lipschitz bounds. Thus, in this section, we develop the needed bracketing results for such Lipschitz classes, for arbitrary domains G. Recall $\mathcal{C}(D, 1, \mathbf{\Gamma}, \mathbf{v})$ is the class of convex functions φ defined on D, uniformly bounded by B and with Lipschitz parameter Γ_i in the direction v_i. When v_i are the standard basis of $\mathbb{R}^d$, we just write $\mathcal{C}(D, 1, \mathbf{\Gamma})$. When we have Lipschitz constraints on convex functions, we will see that the situation for forming brackets for $\mathcal{C}(D, 1, \mathbf{\Gamma})$ with $D \subseteq [0,1]^d$ is essentially the same as for forming brackets for $\mathcal{C}\left([0,1]^d, 1, \mathbf{\Gamma}\right)$. For two sets $C, D \subset \mathbb{R}^d$, define the Hausdorff distance between them by

$$l_H(C,D) := \max\left(\sup_{x\in D}\inf_{y\in C}\|x-y\|, \sup_{y\in C}\inf_{x\in D}\|x-y\|\right)$$

For $B > 0$ and a convex function f defined on a convex set D, define the epigraph $V_B(f)$ by

$$V_B(f) := \{(x_1, \ldots, x_d, x_{d+1}) : (x_1, \ldots, x_d) \in D, f(x_1, \ldots, x_d) \le x_{d+1} \le B\}.$$

Bronshtein (1976) found entropy estimates in the Hausdorff distance for classes of d-dimensional convex sets (see also Dudley (1999), chapter 8). These entropy bounds for classes of convex sets are the main tool for Bronshtein (1976)'s entropy bounds for classes of convex functions, and they will also be the main tool in our bracketing bounds for convex functions (with Lipschitz constraints).

Theorem 2.1 (Bronshtein (1976)). *For any $R > 0$ and any integer $d \geq 1$, there exist positive real numbers c_d and $\epsilon_{0,d}$ such that for all $0 < \epsilon \leq \epsilon_{0,d} R$ there is an ϵ-cover of $\mathcal{K}^{d+1}(R)$ in the Hausdorff distance of cardinality not larger than* $\exp\left\{c_d (R/\epsilon)^{d/2}\right\}$.

The following lemma connects the Hausdorff distance on sets of epigraphs of Lipschitz functions to the supremum distance for those functions.

Lemma 2.1. *Let $G \subseteq [0,1]^d$ be any convex set and $B, \Gamma_1, \ldots, \Gamma_d > 0$. For $f, g \in \mathcal{C}\left(G, B, (\Gamma_1, \ldots, \Gamma_d)\right)$,*

$$\|f - g\|_\infty \leq l_H(V_B(f), V_B(g)) \sqrt{1 + \sum_{i=1}^d \Gamma_i^2}.$$

Proof. For ease of notation, let $\rho = l_H(V_B(f), V_B(g))$. Fix $x \in G$ and suppose $f(x) < g(x)$, without loss of generality. Now, $(x, f(x)) \in V_B(f)$ so there exists $(x', y') \in V_B(g)$ such that $\|(x', y') - (x, f(x))\| \leq \rho$. Since $f(x) < g(x)$, $(x, f(x))$ is outside the epigraph of $V_B(g)$ so by convexity of $V_B(g)$, $y' = g(x')$. Thus

$$0 \leq g(x) - f(x) = g(x) - g(x') + g(x') - f(x) \leq \|x - x'\| \sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2} + |g(x') - f(x)|,$$

since $g(x) - g(x')| = |g(x_1, \ldots, x_d) - g(x_1, \ldots, x_{d-1}, x_d') + \cdots + g(x_1, x_2', \ldots, x_d') - g(x_1', \ldots, x_d')|$ which is bounded above by

$$|x_d - x_d'|\Gamma_d + \cdots + |x_1 - x_1'|\Gamma_1 \leq \|x - x'\| \sqrt{\Gamma_1^2 + \cdots + \Gamma_d^2}$$

by the Cauchy-Schwarz inequality. Thus, again by Cauchy-Schwarz,

$$0 \leq g(x) - f(x) \leq \rho \sqrt{1 + \sum_i \Gamma_i^2},$$

as desired. □

Theorem 3.2 from (Guntuboyina and Sen, 2013) gives the following result when $D = \prod_{i=1}^d [a_i, b_i]$; we now extend it to the case of a general D. When we consider convex functions without Lipschitz constraints, we will partition D into sets that are similar to parallelotopes. Note that if $P \subset R \subset \mathbb{R}^d$ where R is a hyperrectangle and P is a parallelotope defined by vectors $v_1, \ldots, v_d$, then if A is a linear map with $v_1, \ldots, v_d$ as its eigenvectors (thus rescaling P), then AR will not necessarily still be a hyperrectangle, i.e. its axes may no longer be orthogonal. Thus, we cannot argue by simple scaling arguments that bracketing numbers for P scale with the lengths along the vectors v_i.

Theorem 2.2. *Let $a_i < b_i$ and let $D \subset \prod_{i=1}^d [a_i, b_i]$ be a convex set. Let $\mathbf{\Gamma} = (\Gamma_1, \ldots, \Gamma_d)$ and $0 < B, \Gamma_1, \ldots, \Gamma_d < \infty$. Then there exist positive constants $c \equiv c_d$*

and $\epsilon_0 \equiv \epsilon_{0,d}$ *such that*

$$\begin{aligned}\log N_{[\,]}\left(\epsilon \operatorname{Vol}(D)^{1/p}, \mathcal{C}\left(D, B, \boldsymbol{\Gamma}\right), L_p\right) &\leq \log N_{[\,]}\left(\epsilon, \mathcal{C}\left(D, B, \boldsymbol{\Gamma}\right), L_\infty\right) \\ &\leq c\left(\frac{B + \sum_{i=1}^d \Gamma_i(b_i - a_i)}{\epsilon}\right)^{d/2}\end{aligned}$$

for $0 < \epsilon \leq \epsilon_0 \left(B + \sum_{i=1}^d \Gamma_i(b_i - a_i)\right)$ *and* $p \geq 1$.

Proof. The first inequality of the theorem is elementary. We will show the second. First we note the following scaling relationship. For $f \in \mathcal{C}\left(D, B, \boldsymbol{\Gamma}\right)$ we can define $\tilde{f} : \tilde{D} \to \mathbb{R}$, where $\tilde{D} \subseteq [0,1]^d$, by $\tilde{f}(t_1, \ldots, t_d) = f(a_1 + t_1(b_1 - a_1), \ldots, a_d + t_d(b_d - a_d))$. Then $\tilde{f} \in \mathcal{C}\left(\tilde{D}, B, (\Gamma_1(b_1 - a_1), \ldots, \Gamma_d(b_d - a_d))\right)$. This shows that

$$\begin{aligned}&N_{[\,]}\left(\epsilon, \mathcal{C}\left(\tilde{D}, B, (\Gamma_1(b_1 - a_1), \ldots, \Gamma_d(b_d - a_d))\right), L_\infty\right) \\ &\qquad = N_{[\,]}\left(\epsilon, \mathcal{C}\left(D, B, (\Gamma_1, \ldots, \Gamma_d)\right), L_\infty\right).\end{aligned} \tag{2}$$

Thus, we now let $a_i = 0$ and $b_i = 1$ and consider a convex domain $\tilde{D} \subset [0,1]^d$. It is then clear if $f \in \mathcal{C}\left(\tilde{D}, B\right)$ that $V_B(f) \in \mathcal{K}^{d+1}(\sqrt{d + B^2})$, where

$$\mathcal{K}^{d+1}(R) = \{D : D \text{ is a closed, convex set}, D \subseteq B(0, R)\}$$

for $R > 0$. Thus, given an $\left(\epsilon / \left(4\sqrt{1 + \Gamma_1^2 + \cdots + \Gamma_d^2}\right)\right)$-cover in Hausdorff distance of $\mathcal{K}^{d+1}(R)$ of $\tilde{N}$ elements $V_1, \ldots, V_{\tilde{N}}$, we can pick $V_B(f_1), \ldots, V_B(f_N)$ for $N \leq \tilde{N}$, such that $l_H(V_B(f_i), \tilde{V}_i) \leq \epsilon/(4\sqrt{1 + \Gamma_1^2 + \cdots + \Gamma_d^2})$, if such an $f_i \in \mathcal{C}\left(\tilde{D}, B, (\Gamma_1, \ldots, \Gamma_d)\right)$ exists. Then from Lemma 2.1, $[f_i - \epsilon, f_i + \epsilon]$ form an L_∞ bracketing set for $\mathcal{C}\left(\tilde{D}, B, (\Gamma_1, \ldots, \Gamma_d)\right)$. Thus, by Theorem 2.1, for some positive c, ϵ_0,

$$\log N_{[\,]}\left(\epsilon, \mathcal{C}\left(\tilde{D}, B, (\Gamma_1, \ldots, \Gamma_d)\right), L_\infty\right) \leq c\left(\frac{\sqrt{(d + B^2)(1 + \Gamma_1^2 + \cdots + \Gamma_d^2)}}{\epsilon}\right)^{d/2}$$

for $0 < \epsilon \leq \epsilon_0 \sqrt{(d + B^2)(1 + \Gamma_1^2 + \cdots + \Gamma_d^2)}$. Using (2), we see that

$$\log N_{[\,]}\left(\epsilon, \mathcal{C}\left(D, B, (\Gamma_1, \ldots, \Gamma_d)\right), L_\infty\right) \leq c\left(\frac{\sqrt{(d + B^2)(1 + \sum_i \Gamma_i^2(b_i - a_i)^2)}}{\epsilon}\right)^{d/2} \tag{3}$$

for $0 < \epsilon \le \epsilon_0 \sqrt{(d+B^2)\left(1+\sum_{i=1}^d \Gamma_i^2 (b_i - a_i)^2\right)}$. It is immediate that the left side of (3) equals

$$\log N_{[\,]} \left(\frac{\epsilon}{A}, \mathcal{C}\left(D, \frac{B}{A}, \left(\frac{\Gamma_1}{A}, \ldots, \frac{\Gamma_d}{A}\right)\right), L_\infty \right)$$

for any $A > 0$ so that for all $A > 0$ (3) is bounded above by

$$c \left(\frac{\sqrt{(dA^2 + B^2)\left(1 + \sum_i \Gamma_i^2 (b_i - a_i)^2 / A^2\right)}}{\epsilon} \right)^{d/2}$$

for $0 < \epsilon \le \epsilon_0 \sqrt{(dA^2 + B^2)(1 + \sum_i \Gamma_i^2 (b_i - a_i)^2 / A^2)}$. We pick

$$A^2 = \sqrt{\frac{B^2 \sum_{i=1}^d \Gamma_i^2 (b_i - a_i)^2}{d}},$$

which yields

$$\log N_{[\,]} \left(\epsilon, \mathcal{C}\left(D, B, (\Gamma_1, \ldots, \Gamma_d)\right), L_\infty \right) \le c \left(\frac{B + \sqrt{d \sum_i \Gamma_i^2 (b_i - a_i)^2}}{\epsilon} \right)^{d/2}$$

if $0 < \epsilon \le \epsilon_0 \left(B + \sqrt{d \sum_i \Gamma_i^2 (b_i - a_i)^2}\right)$. Since

$$\sqrt{\sum_i \Gamma_i^2 (b_i - a_i)^2} \le \sum_i \Gamma_i (b_i - a_i) \le \sqrt{d \sum_i \Gamma_i^2 (b_i - a_i)^2},$$

which are basic facts about l_p norms in $\mathbb{R}^d$, we are done showing the second inequality of the theorem. $\square$

3 Bracketing without Lipschitz Constraints

In the previous section we bounded bracketing entropy for classes of functions with Lipschitz constraints. In this section we remove those Lipschitz constraints.

3.1 Notation and Assumptions

With Lipschitz constraints we could consider arbitrary domains D, but without the Lipschitz constraints we need more restrictions. We will now require that D is a polytope, and, to begin with, we also assume that D is simple. We will consider only the case $d \ge 2$ since the result is given when $d = 1$ in Dryanov (2009).

Assumption 1. *Let $d \ge 2$ and let $D \subset \mathbb{R}^d$ be a simple convex polytope, meaning that all $(d-k)$-dimensional faces of D have exactly k incident facets.*

It is well-known that the simplicial polytopes are dense in the class of all polytopes in the Hausdorff distance. The simple polytopes are dual to the simplicial ones, and are also dense in the class of all polytopes in the Hausdorff distance (page 82 of Grünbaum (1967)). Any convex polytope can be triangulated into $O(n^{\lfloor d/2 \rfloor})$ simplices (which are simple polytopes) if the polytope has n vertices, see e.g. Dey and Pach (1998), and so one can translate our results to a general polytope D. However, then any geometric intuition provided by the constants in the bounds is lost.

We let $D = \cap_{j=1}^N E_j$ where $E_j := \left\{x \in \mathbb{R}^d : \langle v_j, x\rangle \geq p_j\right\}$ are halfspaces with (inner) normal vectors v_j, and where $p_j \in \mathbb{R}$, for $j = 1, \ldots, N$. Let $H_j := \left\{x \in \mathbb{R}^d : \langle x, v_j\rangle = p_j\right\}$ be the corresponding hyperplanes. For $k \in \mathbb{N}$, let $J_k := \left\{(j_1, \ldots, j_k) \in \{1, \ldots, N\}^k : j_1 < \cdots < j_k\right\}$ and $I_k := \{0, \ldots, A\}^k$. For $\boldsymbol{j} \in J_k$, let

$$G_{\boldsymbol{j}} = \cap_{\alpha=1}^k H_{j_\alpha}.$$

Any $G_{\boldsymbol{j}}$, $\boldsymbol{j} \in J_k$, is $(d-k)$-dimensional and so, by Fritz John's theorem (John (1948), see also Ball (1992) or Ball (1997)), contains a $(d-k)$-dimensional ellipsoid $A_{\boldsymbol{j}} - x_{\boldsymbol{j}}$ of maximal $(d-k)$-dimensional volume, such that

$$A_{\boldsymbol{j}} - x_{\boldsymbol{j}} \subset G_{\boldsymbol{j}} - x_{\boldsymbol{j}} \subset d(A_{\boldsymbol{j}} - x_{\boldsymbol{j}}) \tag{4}$$

for some point $x_{\boldsymbol{j}} \in G_{\boldsymbol{j}}$. Let $e_{k+1}, \ldots, e_d$ be the orthonormal basis given by the axes of the ellipsoid $A_{\boldsymbol{j}} - x_{\boldsymbol{j}}$ and let $\gamma_{\boldsymbol{j},\alpha}/2$ be the radius of $A_{\boldsymbol{j}}$ in the direction e_α, meaning that $x_{\boldsymbol{j}} \pm \gamma_{\boldsymbol{j},\alpha} e_\alpha/2$ lies in the boundary of $A_{\boldsymbol{j}}$. We will rely heavily on Fritz John's theorem to understand the size of $G_{\boldsymbol{j}}$. Let $d^+(x, \partial G_{\boldsymbol{j}}, e) := \inf_{K>0} \{K : x + Ke \cap \partial G_{\boldsymbol{j}} \neq \emptyset\}$ and let

$$u := 2^{-2(p+1)^2(p+2)} \wedge \min_{k \in \{1,\ldots,d-1\}} \min_{\boldsymbol{j} \in J_k, e \in \operatorname{span}\{e_{k+1},\ldots,e_d\}} \frac{d^+(x_{\boldsymbol{j}}, \partial G_{\boldsymbol{j}}, e)}{L_{k,2}} \tag{5}$$

where

$$L_{k,2} := 1 \vee \sup_{\beta > k} \sum_{\gamma=1}^k \frac{\left\langle \tilde{f}_\gamma, v_{j_\beta}\right\rangle}{\left\langle \tilde{f}_\gamma, v_{j_\gamma}\right\rangle} \tag{6}$$

and $\tilde{f}_\gamma$ are defined in Proposition 4.2. Then let

$$0 = \delta_0 < \delta_1 < \cdots < \delta_A < u = \delta_{A+1} < \delta_{A+2} = \infty \tag{7}$$

be a sequence to be defined later.

Let $\operatorname{Lin} P$ be the translated affine span of P, i.e. the space of all linear combinations of elements of $(P - x)$, for any $x \in P$. Note that $\operatorname{lin} P$ is commonly used to refer to the linear span of P rather than of $P - x$, and thus to distinguish from this case, we use the notation "Lin" rather than "lin." For a point x, a set H, and a unit vector v, let

$$d(x, H, v) := \inf \{|k| : x + kv \in H\}$$

be the distance from x to H in direction v, and for a set E, $d(E, H, v) := \inf_{x \in E} d(x, H, v)$. For $\boldsymbol{i} = (i_1, \ldots, i_k) \in I_k$ and $\boldsymbol{j} = (j_1, \ldots, j_k) \in J_k$ let

$$G_{\boldsymbol{i},\boldsymbol{j}} := \{x \in D : \delta_{i_\alpha} \le d(x, H_{j_\alpha}) < \delta_{i_\alpha+1} \text{ for } \alpha = 1, \ldots, N\}, \tag{8}$$

where for $\alpha > k$ we let $i_\alpha = A + 1$. These sets are not parallelotopes, since for $\alpha > k$, $\delta_{i_\alpha+1} = \infty$. However, for any $x \in G_{\boldsymbol{j}}$, $(G_{\boldsymbol{i},\boldsymbol{j}} - x) \cap \operatorname{span}\{v_{j_1}, \ldots, v_{j_\beta}\}$, for $\beta \le k$, is contained in a β-dimensional parallelotope.

3.2 Main Results

We want to bound the slope of functions $f \in \mathcal{C}(D, 1)|_{G_{\boldsymbol{i},\boldsymbol{j}}}$, so that we can apply bracketing bounds on convex function classes with Lipschitz bounds. Note that each $G_{\boldsymbol{i},\boldsymbol{j}}$ is distance δ_{i_α} in the direction of v_{j_α} from H_{j_α}, which means that if $f \in \mathcal{C}(D, 1)|_{G_{\boldsymbol{i},\boldsymbol{j}}}$ then f has Lipschitz constant bounded by $2/\delta_{i_\alpha}$ along the direction v_{j_α} towards H_{j_α}. However, the vectors v_{j_α} are not orthonormal, so the distance from $G_{\boldsymbol{i},\boldsymbol{j}}$ along v_{j_α} to a hyperplane other than H_{j_α} may be smaller than δ_{i_α}.

For each $G_{\boldsymbol{i},\boldsymbol{j}}$ we will find an orthonormal basis such that $G_{\boldsymbol{i},\boldsymbol{j}}$ is contained in a rectangle R whose axes are given by the basis and whose lengths along those axes (i.e., widths) is bounded by a constant times the width of one of the normal vectors v_{j_α}. Furthermore, the distance from R along each basis vector to ∂D will be bounded by the distance from $G_{\boldsymbol{i},\boldsymbol{j}}$ along v_{j_α} to H_{j_α}. This will give us control of both the Lipschitz parameters and the widths corresponding to the basis, and thus control of the size of bracketing for classes of convex functions.

Proposition 3.1. *Let Assumption 1 hold for a convex polytope D. For each $k \in \{0, \ldots, d\}$, $\boldsymbol{i} \in I_k, \boldsymbol{j} \in J_k$, and each $G_{\boldsymbol{i},\boldsymbol{j}}$, there is an orthornormal basis $\mathbf{e}_{\boldsymbol{i},\boldsymbol{j}} \equiv \mathbf{e} := (e_1, \ldots, e_d)$ of $\mathbb{R}^d$ such that for any $f \in \mathcal{C}(D, B)|_{G_{\boldsymbol{i},\boldsymbol{j}}}$, f has Lipschitz constant $2B/\delta_{i_\alpha}$ in the direction e_α, where $\delta_{i_\alpha} = \delta_{A+1}$ if $k+1 \le \alpha \le d$. Furthermore, for $\alpha = 1, \ldots, k$, $e_{\boldsymbol{i},\boldsymbol{j},\alpha} \equiv e_\alpha$ satisfies*

$$e_\alpha \in \operatorname{span}\{v_{j_1}, \ldots, v_{j_\alpha}\}, \; e_\alpha \perp \operatorname{span}\{v_{j_1}, \ldots, v_{j_{\alpha-1}}\}, \quad \text{and } \langle e_\alpha, v_\alpha \rangle > 0,$$

and for $\alpha \in \{k+1, \ldots, d\}$, $e_\alpha \perp \operatorname{span}\{v_{j_1}, \ldots, v_{j_k}\}$.

Proof. Without loss of generality, for ease of notation we assume in this proof that

$$j_\beta = \beta \text{ for } \beta = 1, \ldots, k,$$

and then that

$$\delta_{i_1} \le \delta_{i_2} \le \cdots \le \delta_{i_k} \le \delta_{i_{k+1}} = \cdots = \delta_{i_N},$$

where we let $i_\alpha = A + 1$ for $k < \alpha \le N$. That is, we assume that $H_1, \ldots, H_k$ are the nearest hyperplanes to $G_{\boldsymbol{i},\boldsymbol{j}}$, in order of increasing distance. To define the orthonormal basis vectors, we will use a Gram-Schmidt orthonormalization, proceeding according to increasing distances from $G_{\boldsymbol{i},\boldsymbol{j}}$ to the hyperplanes H_j. Define $e_1 := v_1$ and for $1 < j \le k$, define e_j inductively by

$$e_j \in \operatorname{span}\{v_1, \ldots, v_j\}, \; e_j \perp \operatorname{span}\{v_1, \ldots, v_{j-1}\}, \; \langle e_j, v_j \rangle > 0, \text{ and } \|e_j\| = 1,$$

and let $\{e_j\}_{j=k+1}^d$ be any orthonormal basis of $\operatorname{span}\{v_1,\ldots,v_k\}^\perp$.

For $\alpha \in \{1,\ldots,k\}$, for any $x \in G_{\boldsymbol{i},\boldsymbol{j}}$, since $d(x,H_\alpha,v)$ is smallest when v is v_α,

$$\begin{aligned}
&d(x,H_\alpha,e_\alpha) \geq d(x,H_\alpha,v_\alpha) \geq \delta_{i_\alpha},\\
&d(x,H_j,e_\alpha) \geq d(x,H_j,v_j) \geq \delta_{i_j} \geq \delta_{i_\alpha}, \text{ for all } N \geq j > \alpha, \text{ and}\\
&d(x,H_j,e_\alpha) = \infty > \delta_{i_\alpha} \text{ for } j < \alpha,
\end{aligned}$$

since $e_\alpha \perp \operatorname{span}\{v_1,\ldots,v_{\alpha-1}\}$. Similarly, for $\alpha \in \{k+1,\ldots,d\}$,

$$\begin{aligned}
&d(x,H_j,e_\alpha) \geq d(x,H_j,v_j) \geq \delta_{A+1}, \text{ for all } N \geq j \geq k+1, \text{ and}\\
&d(x,H_j,e_\alpha) = \infty > \delta_{A+1} \text{ for } j \leq k,
\end{aligned}$$

since $e_\alpha \perp \operatorname{span}\{v_1,\ldots,v_k\}$. Thus, we have $d(G_{\boldsymbol{i},\boldsymbol{j}},H_j,e_\alpha) \geq \delta_{i_\alpha}$ for $\alpha \in \{1,\ldots,d\}$ and for $j \in \{1,\ldots,N\}$. That is, we have shown

$$d(G_{\boldsymbol{i},\boldsymbol{j}},\partial D,e_\alpha) \geq \delta_{i_\alpha} \text{ for all } \alpha \in \{1,\ldots,d\}\,. \tag{9}$$

Thus, if $f \in \mathcal{C}(D,B)|_{G_{\boldsymbol{i},\boldsymbol{j}}}$, then for any $x \in G_{\boldsymbol{i},\boldsymbol{j}}$, let $z_1 = x - \gamma_1 e_\alpha$ and $z_2 = x + \gamma_2 e_\alpha$, $\gamma_1, \gamma_2 > 0$, both be elements of $\partial G_{\boldsymbol{i},\boldsymbol{j}}$, so that by convexity we have

$$\frac{-2B}{\delta_{i_\alpha}} \leq \frac{f(z_1) - f(z_1 - \delta_{i_\alpha} e_\alpha)}{\delta_{i_\alpha}} \leq \frac{f(x+ke_\alpha) - f(x)}{k} \leq \frac{f(z_2 + \delta_{i_\alpha} e_\alpha) - f(z_2)}{\delta_{i_\alpha}} \leq \frac{2B}{\delta_{i_\alpha}},$$

using (9). Thus, f satisfies a Lipschitz constraint in the direction of e_α. □

Here is our main theorem. It gives a bracketing entropy of $\epsilon^{-d/2}$ when D is a fixed simple polytope. Its proof relies on embedding $G_{\boldsymbol{i},\boldsymbol{j}}$ in a rectangle $R_{\boldsymbol{i},\boldsymbol{j}}$ with axes given by Proposition 3.1. We need to control the distance of $G_{\boldsymbol{i},\boldsymbol{j}}$ to ∂D, and we need to control the size of $R_{\boldsymbol{i},\boldsymbol{j}}$ in terms of the widths along its axes. Then we can use the results of Section 2 on $R_{\boldsymbol{i},\boldsymbol{j}}$ and thus on $G_{\boldsymbol{i},\boldsymbol{j}}$. Our studying the size of $R_{\boldsymbol{i},\boldsymbol{j}}$ is somewhat lengthy so we defer that until Section 4. The constant S has an explicit form given in the proof of the theorem.

Theorem 3.1. *Let Assumption 1 hold for a convex polytope $D \subseteq \prod_{i=1}^d [a_i,b_i]$, for an integer $d \geq 2$. Fix $p \geq 1$. Then for some $\epsilon_0 > 0$ and for $0 < \epsilon \leq \epsilon_0 B \left(\prod_{i=1}^d b_i - a_i\right)^{1/p}$,*

$$\log N_{[\,]}\left(\epsilon, \mathcal{C}(D,B), L_p\right) \leq S \left(\frac{B\left(\prod_{i=1}^d (b_i - a_i)\right)^{1/p}}{\epsilon} \right)^{d/2},$$

where S is a constant depending on d and D (and on u, which is fixed by (5)).

Proof. First, we will reduce to the case where $D \subset [0,1]^d$ and $B = 1$ by a scaling argument. Let A be an affine map from $\prod_{i=1}^d [a_i, b_i]$ to $[0,1]$, where $\tilde{D}$ is the image of D, and assume we have a bracketing cover $[\tilde{l}_1, \tilde{u}_1], \dots, [\tilde{l}_N, \tilde{u}_N]$ of $\mathcal{C}\left(\tilde{D}, 1\right)$. Let $l_i := B\tilde{l}_i \circ A$ and similarly for u_i, so that $[l_1, u_1], \dots, [l_N, u_N]$ form brackets for $\mathcal{C}(D, B)$. Their L_p^p size is

$$\int_D (u_i(x) - l_i(x))^p \, dx = B^p \int_{\tilde{D}} (\tilde{u}_i(x) - \tilde{l}_i(x))^p \prod^d (b_i - a_i) dx.$$

Thus,

$$N_{[]}\left(\epsilon B \left(\prod^d b_i - a_i\right)^{1/p}, \mathcal{C}(D, B), L_p\right) \le N_{[]}\left(\epsilon, \mathcal{C}\left(\tilde{D}, 1\right), L_p\right),$$

so apply the theorem with $\eta = \epsilon / B \left(\prod^d b_i - a_i\right)^{1/p}$ for ϵ. Note that the constant S depends on $\tilde{D}$, the version of D normalized to lie in $[0,1]^d$.

We now assume $D \subset [0,1]^d$ and $B = 1$. For a sequence $a_{\boldsymbol{i},k} > 0$ (constant over $\boldsymbol{j} \in J_k$), to be defined later, let $a = \left(\sum_{k=0}^d \sum_{\boldsymbol{j} \in J_k, \boldsymbol{i} \in I_k} a_{\boldsymbol{i},k}^p \operatorname{Vol}_d(G_{\boldsymbol{i},\boldsymbol{j}})\right)^{1/p}$. By Assumption 1 and since $D \subset \cup_{k=0}^d \cup_{\boldsymbol{j} \in J_k, \boldsymbol{i} \in I_k} G_{\boldsymbol{i},\boldsymbol{j}}$,

$$N_{[]}(a, \mathcal{C}(D, 1), L_p) \le \prod_{k=0}^d \prod_{\boldsymbol{j} \in J_k, \boldsymbol{i} \in I_k} N_{[]}\left(a_{\boldsymbol{i},k} \operatorname{Vol}_{d-k}(G_{\boldsymbol{i},\boldsymbol{j}})^{1/p}, \mathcal{C}(D, 1)|_{G_{\boldsymbol{i},\boldsymbol{j}}}, L_p\right),$$

as in (1). Now by Lemma 4.3, we can ignore all terms with $\boldsymbol{j} \in J_k \setminus J_k^D$, where $J_k^D := \left\{\boldsymbol{j} \in J_k : \cap_{\alpha=1}^k H_{j_\alpha} \text{ is a } k\text{-face of } G\right\}$. Thus

$$\log N_{[]}(a, \mathcal{C}(D, 1), L_p) \le \sum_{k=0}^d \sum_{\boldsymbol{j} \in J_k^D} \sum_{\boldsymbol{i} \in I_k} \log N_{[]}\left(a_{\boldsymbol{i},k} \operatorname{Vol}_{d-k}(G_{\boldsymbol{i},\boldsymbol{j}})^{1/p}, \mathcal{C}(D, 1)|_{G_{\boldsymbol{i},\boldsymbol{j}}}, L_p\right).$$

First we compute the sum over I_k for a fixed $\boldsymbol{j} \in J_k$. Thus by Proposition 3.1,

$$\mathcal{C}(D, 1)|_{G_{\boldsymbol{i},\boldsymbol{j}}} \subset \mathcal{C}(G_{\boldsymbol{i},\boldsymbol{j}}, 1, \boldsymbol{\Gamma}, \mathbf{e}) \tag{10}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{i}} = (2/\delta_{i_1}, \dots, 2/\delta_{i_k}, 2/u, \dots, 2/u)$. Let $R_{\boldsymbol{i},\boldsymbol{j}}$ be as in (25). That is, let $\rho_{\boldsymbol{j},\alpha} = w(G_{\boldsymbol{j}}, e_\alpha)$, $L_{k,1}$ be given by (22), and let

$$R_{\boldsymbol{i},\boldsymbol{j}} := \sum_{\alpha=1}^k [\alpha!(\delta_{i_\alpha+1} - \delta_{i_\alpha}) e_\alpha, \alpha!(\delta_{i_\alpha+1} - \delta_{i_\alpha}) e_\alpha] + \sum_{\alpha=k+1}^d [-2L_{k,1}\rho_{\boldsymbol{j},\alpha} e_\alpha, 2L_{k,1}\rho_{\boldsymbol{j},\alpha} e_\alpha],$$

so that $G_{\boldsymbol{i},\boldsymbol{j}} \subseteq x + R_{\boldsymbol{i},\boldsymbol{j}}$ for any $x \in G_{\boldsymbol{i},\boldsymbol{j}}$ by (26). Then by (10) (and the first inequality of Theorem 2.2) we bound

$$\sum_{\boldsymbol{i} \in I_k} \log N_{[]}\left(a_{\boldsymbol{i},k} \operatorname{Vol}(G_{\boldsymbol{i},\boldsymbol{j}})^{1/p}, \mathcal{C}(D, 1)|_{G_{\boldsymbol{i},\boldsymbol{j}}}, L_p\right) \le \sum_{\boldsymbol{i} \in I_k} \log N_{[]}(a_{\boldsymbol{i},k}, \mathcal{C}(G_{\boldsymbol{i},\boldsymbol{j}}, 1, \boldsymbol{\Gamma}_{\boldsymbol{i}}), L_\infty). \tag{11}$$

We use the trivial bracket $[-1,1]$ for any $G_{\boldsymbol{i},\boldsymbol{j}}$ where $i_\alpha = 0$ for any $\alpha \in \{1,\ldots,k\}$, and otherwise we use Theorem 2.2, which shows us that (11) is bounded by

$$\sum_{i_1=1}^{A} \cdots \sum_{i_k=1}^{A} c \left(\frac{1 + \sum_{\alpha=1}^{k} \frac{2d!(\delta_{i_\alpha+1} - \delta_{i_\alpha})}{\delta_{i_\alpha}} + \sum_{\alpha=k+1}^{d} \frac{8L_{k,1}\rho_{\boldsymbol{j},\alpha}}{u}}{a_{\boldsymbol{i},k}} \right)^{d/2}. \tag{12}$$

For $\boldsymbol{i} \in I_k$, we will let $a_{(i_1,\ldots,i_k)} = 1$ if $i_\alpha = 0$ for any $\alpha \in \{1,\ldots,k\}$, and otherwise we let

$$a_{(i_1,\ldots,i_k)} := \prod_{\beta=1}^{k} a_{i_\beta} := \prod_{\beta=1}^{k} \epsilon^{1/k} \exp\left\{ -p \frac{(p+1)^{i_\beta - 2}}{(p+2)^{i_\beta - 1}} \log \epsilon \right\}, \text{ and}$$
$$\delta_i := \exp\left\{ p \left(\frac{p+1}{p+2} \right)^{i-1} \log \epsilon \right\} \quad \text{for } i = 1,\ldots,A,$$

and $\delta_0 = 0$. Since $L_{k,1} \geq 1$, $L_{k,2} \geq 1$, and $u \leq \rho_{\boldsymbol{j},\alpha}/L_{k,2}$ by (5) for all $k, \boldsymbol{i}, \boldsymbol{j}$ and $\alpha = k+1,\ldots,d$, we have $\sum_{\alpha=k+1}^{d} \frac{8L_{k,1}\rho_{\boldsymbol{j},\alpha}}{u} \leq \prod_{\alpha=k+1}^{d} \frac{8L_{k,1}\rho_{\boldsymbol{j},\alpha}}{u}$ (using the fact that for $a, b \geq 2$, $ab \geq a+b$). Similarly, $\sum_{\alpha=1}^{k} 2(\delta_{i_\alpha+1} - \delta_{i_\alpha})/\delta_{i_\alpha} \leq \prod_{\alpha=1}^{k} 2\delta_{i_\alpha+1}/\delta_{i_\alpha}$ since $2\delta_{i_\alpha+1}/\delta_{i_\alpha} > 2$. Thus (12) is bounded above by

$$c(d!)^{d/2} \left(1 + \prod_{\alpha=k+1}^{d} \frac{8L_{k,1}\rho_{\boldsymbol{j},\alpha}}{u} \right)^{d/2} \sum_{i_1=1}^{A} \cdots \sum_{i_k=1}^{A} a_{\boldsymbol{i},k}^{-d/2} \prod_{\alpha=1}^{k} \left(\frac{2\delta_{i_\alpha+1}}{\delta_{i_\alpha}} \right)^{d/2}, \tag{13}$$

which is

$$c(d!)^{d/2} \left(1 + \prod_{\alpha=k+1}^{d} \frac{8L_{k,1}\rho_{\boldsymbol{j}\alpha}}{u} \right)^{d/2} \sum_{i_1=1}^{A} \cdots \sum_{i_k=1}^{A} \prod_{\beta=1}^{k} \left(\frac{2\delta_{i_\beta+1}}{\delta_{i_\beta} a_{i_\beta}} \right)^{d/2}. \tag{14}$$

For $i = 1,\ldots,A$, let $\zeta_i := \sqrt{\epsilon^{1/k}\delta_{i+1}/(\delta_i a_i)}$, so that $\sum_{i_1=1}^{A} \cdots \sum_{i_k=1}^{A} \prod_{\beta=1}^{k} \left(\frac{2\delta_{i_\beta+1}}{\delta_{i_\beta} a_{i_\beta}} \right)^{d/2}$ equals

$$\begin{aligned} \sum_{i_1=1}^{A} \cdots \sum_{i_k=1}^{A} 2^{kd/2} \epsilon^{-d/2} \prod_{\beta=1}^{k} \zeta_{i_\beta}^{d} &= 2^{kd/2} \epsilon^{-d/2} \sum_{i_1=1}^{A} \zeta_{i_1}^{d} \sum_{i_2=1}^{A} \zeta_{i_2}^{d} \cdots \sum_{i_k=1}^{A} \zeta_{i_k}^{d} \\ &= \epsilon^{-d/2} 2^{kd/2} B_u^k \end{aligned}$$

where

$$B_u := \sum_{i=1}^{A} \zeta_i^d \leq 2u^{d/(2(p+1)^2)}, \tag{15}$$

by Lemma (3.1).

Next, we will relate the term $c(d!)^{d/2}\left(1+\prod_{\alpha=k+1}^{d}\frac{8L_{k,1}\rho_{j\alpha}}{u}\right)^{d/2}$ to $\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})$. Recall $A_{\boldsymbol{j}}$ is the ellipsoid defined in (4) which has diameter (and width) in the e_α direction given by $\gamma_{\boldsymbol{j},\alpha}$. By (4),

$$\rho_{\boldsymbol{j},\alpha} \le d\gamma_{\boldsymbol{j},\alpha}.$$

The volume of $A_{\boldsymbol{j}}$ is $\mathrm{Vol}_{d-k}(A_{\boldsymbol{j}}) = \left(\prod_{\alpha=k+1}^{d}\gamma_{\boldsymbol{j},\alpha}/2\right)\pi^{(d-k)/2}/\Gamma((d-k)/2+1)$. Thus, letting $C_d := \frac{(2d)^{d-k}\Gamma((d-k)/2+1)}{\pi^{(d-k)/2}}$, we have

$$\prod_{\alpha=k+1}^{d}\rho_{j,\alpha} \le C_d\,\mathrm{Vol}_{d-k}(A_{\boldsymbol{j}}) \le C_d\,\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}}).$$

Then we have shown that (14) is bounded above by

$$c_d(d!)^{d/2}2^{kd/2}\left(1+\left(\frac{8}{u}\right)^{d-k}C_d\,\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})\prod_{\alpha=k+1}^{d}L_{k,1}\right)^{d/2}B_u^k\cdot\epsilon^{-d/2}. \tag{16}$$

Then, gathering the constants together into $\tilde{c}_d$, we have shown

$$\sum_{\boldsymbol{i}\in I_k}\log N_{[\,]}\left(a_{\boldsymbol{i},k}\,\mathrm{Vol}(G_{\boldsymbol{i},\boldsymbol{j}})^{1/p},\mathcal{C}\left(D,1\right)|_{G_{\boldsymbol{i},\boldsymbol{j}}},L_p\right)$$
$$\le \epsilon^{-d/2}\tilde{c}_d\left(\frac{\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})\prod_{\alpha=k+1}^{d}L_{k,1}}{u^{d-k}}\right)^{d/2}u^{kd/(2(p+1)^2)}.$$

Then the cardinality of the collection of brackets covering the entire domain D is given by summing over $\boldsymbol{j}\in J_k$ and $k\in\{0,\ldots,d\}$.

We have computed the cardinality of the brackets. Now we bound their size. We have

$$a^p \le \sum_{k=0}^{d}(2L_{k,1})^{d-k}\sum_{\boldsymbol{j}\in J_k}\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})\sum_{\boldsymbol{i}\in I_k}a_{\boldsymbol{i},k}^p\prod_{\alpha=1}^{k}\frac{\delta_{i_\alpha+1}-\delta_{i_\alpha}}{\left\langle\tilde{f}_\alpha,v_{j_\alpha}\right\rangle} \tag{17}$$

by Proposition 4.2, with $\tilde{f}_\alpha$ defined there. Fixing k, we have

$$\sum_{\boldsymbol{j}\in J_k}\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})\sum_{\boldsymbol{i}\in I_k}a_{\boldsymbol{i},k}^p\prod_{\alpha=1}^{k}\frac{\delta_{i_\alpha+1}-\delta_{i_\alpha}}{\left\langle\tilde{f}_\alpha,v_{j_\alpha}\right\rangle} \le \sum_{\boldsymbol{j}\in J_k}\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})L_{\boldsymbol{j},3}^k\sum_{i_1=0}^{A}\cdots\sum_{i_k=0}^{A}\prod_{\alpha=1}^{k}a_{i_\alpha}^p\delta_{i_\alpha+1}$$
$$\le \sum_{\boldsymbol{j}\in J_k}\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}})L_{\boldsymbol{j},3}^k\sum_{i_1=0}^{A}a_{i_1}^p\delta_{i_1+1}\cdots\sum_{i_k=0}^{A}a_{i_k}^p\delta_{i_k+1}.$$

where $L_{\boldsymbol{j},3} := \max_{\alpha\in\{1,\ldots,k\}}1/\left\langle\tilde{f}_\alpha,v_{j_\alpha}\right\rangle$. We have

$$\sum_{\alpha=0}^{A}a_\alpha^p\delta_{\alpha+1} = \epsilon^{p/k}\left(1+\sum_{\alpha=1}^{A}\zeta_\alpha^2\right) =: \epsilon^{p/k}A_u, \tag{18}$$

where $A_u \le 1 + 2u^{1/(p+1)^2}$ by Lemma 3.1. Thus

$$\sum_{\boldsymbol{j}\in J_k} \mathrm{Vol}_{d-k}(G_{\boldsymbol{j}}) L_{\boldsymbol{j},3}^k \left(\sum_{i_1=0}^{A} a_{i_1}^p \delta_{i_1+1} \right) \cdots \left(\sum_{i_k=0}^{A} a_{i_k}^p \delta_{i_k+1} \right) \le \epsilon^p A_u^k \sum_{\boldsymbol{j}\in J_k} \mathrm{Vol}_{d-k}(G_{\boldsymbol{j}}) L_{\boldsymbol{j},3}^k,$$

so by (17)

$$a \le \epsilon \left(\sum_{k=0}^{d} (2L_{k,1})^{d-k} S_k^D A_u^k \right)^{1/p}$$

where $S_k^D = \sum_{j\in J_k} \mathrm{Vol}_{d-k}(G_{\boldsymbol{j}}) L_{\boldsymbol{j},3}^k$. □

Lemma 3.1. *For any $\gamma \ge 1$, with A and u given by (7),*

$$\sum_{\alpha=1}^{A} \zeta_\alpha^\gamma \le 2u^{\gamma/(2(p+1)^2)}.$$

Proof. Taking $\epsilon \le \epsilon_0 \le 1$, $\zeta_\alpha \le 1$. Then for $\alpha = 2, \ldots, A$,

$$\begin{aligned} \frac{\zeta_\alpha}{\zeta_{\alpha+1}} &= \exp\left\{ \frac{-p\log\epsilon}{2(p+1)^2(p+2)} \left(\frac{p+1}{p+2}\right)^{\alpha-1} \right\} \\ &\ge \exp\left\{ \frac{-p\log\epsilon}{2(p+1)^2(p+2)} \left(\frac{p+1}{p+2}\right)^{A-1} \right\} \\ &\ge \exp\left\{ \frac{-\log u}{2(p+1)^2(p+2)} \right\} =: R. \end{aligned}$$

Then, $\zeta_\alpha^\gamma(R^\gamma - 1) \le \zeta_\alpha^\gamma R^\gamma - (R\zeta_{\alpha-1})^\gamma$ so $\zeta_\alpha^\gamma \le (R^\gamma/(R^\gamma-1))(\zeta_\alpha^\gamma - \zeta_{\alpha-1}^\gamma)$ and thus

$$\sum_{\alpha=1}^{A} \zeta_\alpha^\gamma \le \zeta_1^\gamma + \frac{R^\gamma}{R^\gamma-1} \sum_{\alpha=2}^{A} (\zeta_\alpha^\gamma - \zeta_{\alpha-1}^\gamma) = \zeta_1^\gamma + \frac{R^\gamma}{R^\gamma-1}(\zeta_A^\gamma - \zeta_1^\gamma) \le \frac{R^\gamma}{R^\gamma-1}\zeta_A^\gamma$$

and $\zeta_A^\gamma = u^{\gamma/(2(p+1)^2)}$. Since $u \le \exp\left(-2(p+1)^2(p+2)\log 2\right)$ by its definition (5), $R \ge 2$ so $R^\gamma/(R^\gamma - 1) \le 2$ for any $\gamma \ge 1$. □

Since simplices are simple polytopes, by triangulating any convex polytope D into simplices, we can extend our theorem to any polytope D. The constant in the bound then depends on the triangulation of D.

Corollary 3.1. *Fix $d \ge 1$ and $p \ge 1$. Let $D \subseteq \prod_{i=1}^d [a_i, b_i]$ be any convex polytope. Then for some $\epsilon_0 > 0$ and for $0 < \epsilon \le \epsilon_0 B \left(\prod_{i=1}^d b_i - a_i\right)^{1/p}$,*

$$\log N_{[\,]}\left(\epsilon, \mathcal{C}(D, B), L_p\right) \lesssim \left(\frac{B\left(\prod_{i=1}^d (b_i - a_i)\right)^{1/p}}{\epsilon} \right)^{d/2}.$$

Proof. By the same scaling argument as in the proof of Theorem 3.1 we may assume $[a_i, b_i] = [0, 1]$ and $B = 1$. The $d = 1$ case is given by Dryanov (2009). Any convex polytope D can be triangulated into d-dimensional simplices (see e.g. Dey and Pach (1998), Rothschild and Straus (1985)). We are done by applying Theorem 3.1 to each of those simplices, by (1). □

4 Proofs: Relating $G_{\boldsymbol{i},\boldsymbol{j}}$ to a Hyperrectangle

4.1 Inscribing $G_{\boldsymbol{i},\boldsymbol{j}}$ in a Hyperrectangle

Theorem 2.2 shows that the bracketing entropy of $\mathcal{C}(D, B, \boldsymbol{\Gamma})$ depends on the diameters of the hyperrectangle $\prod_{i=1}^d [a_i, b_i]$ circumscribing D. This is part of why bounding entropies on hyperrectangular domains is more straightforward than on non-hyperrectangular domains. In this section we prove Propositions 4.1 and 4.2, which show how to embed the domains $G_{\boldsymbol{i},\boldsymbol{j}}$, which partition D, into hyperrectangles. We used this in the proof of Theorem 3.1 so we could apply Theorem 2.2.

The support function for a convex set D is, for $x \in \mathbb{R}^d$,

$$h(D, x) := \max_{d \in D} \langle d, x \rangle .$$

Then the width function is, for $\|u\| = 1$,

$$w(D, u) := h(D, u) + h(D, -u),$$

which gives the distance between supporting hyperplanes of D with inner normal vectors u and $-u$, respectively, and let

$$w(D) = \sup_{\|u\|=1} w(D, u).$$

Theorem 2.2 says that the bracketing entropy of convex functions on domain D with Lipschitz constraints along directions $e_1, \ldots, e_k$ depends on $w(D, e_i)$ (since that gives the maximum "rise" in "rise over run"). In our proof of Theorem 3.1 we partitioned D into sets related to parallelotopes. Thus we will study the widths of parallelotopes. We know the width of $G_{\boldsymbol{i},\boldsymbol{j}}$ in the directions v_{j_α}, which are $\delta_{i_\alpha+1} - \delta_{i_\alpha}$, by definition.

Lemma 4.1. *Let V be a vector space of dimension $j \in \mathbb{N}$ containing linearly independent vectors $v_1, \ldots, v_j$. Let $d_i > 0$ for $i = 1, \ldots, j$, and let P be the parallelotope defined by having $w(P, v_i) = d_i$. Then P satisfies*

$$w(P) \le j! \max_{1 \le i \le j} d_i.$$

Proof. The proof is by induction. The case $j = 1$ is trivial. Now assume the statement holds for $j - 1$ and we want to show it for j. For any $x, y \in \partial P$ we can find a path $x = x_0, x_1, \ldots, x_n = y$ from x to y such that x_i and x_{i+1} are elements

of (the boundary of) the same facet of P. P has $2j$ facets; if $n > j$, then we can find a path through the complementary $2j - n$ facets, so that we may assume $n \le j$. By the induction hypothesis, $\|x_{i+1} - x_i\| \le (j-1)! \max_{1 \le i \le j} d_i$, since any $(j-1)$-dimensional facet is a parallelotope lying in a hyperplane with normal vector v_i and widths $d_1, \ldots, d_{i-1}, d_{i+1}, \ldots d_j$. Thus

$$\|x - y\| \le \sum_{i=1}^{n} \|x_i - x_{i-1}\| \le n(j-1)! \max_{1 \le i \le j} d_i \le j! \max_{1 \le i \le j} d_i,$$

as desired. ☐

This gives a bound on the width of $G_{\boldsymbol{i},\boldsymbol{j}}$ in the direction of each basis vector e_α, $\alpha = 1, \ldots, k$, from Proposition 3.1.

Proposition 4.1. *Let Assumption 1 hold for a convex polytope D. Fix $k \in \{0, \ldots, d\}$, $\boldsymbol{i} \in I_k, \boldsymbol{j} \in J_k$, and let $G_{\boldsymbol{i},\boldsymbol{j}}$ be as in (8). Let $\mathbf{e}_{\boldsymbol{i},\boldsymbol{j}} \equiv \mathbf{e} := (e_1, \ldots, e_d)$, with $e_\alpha \in \mathbb{R}^d$, be the orthornormal basis from Proposition 3.1. Then*

$$w(G_{\boldsymbol{i},\boldsymbol{j}}, e_\alpha) \le \alpha!(\delta_{i_\alpha+1} - \delta_{i_\alpha}).$$

for $\alpha = 1, \ldots, k$.

Proof. Let $\alpha \in \{1, \ldots, k\}$ and let $w(G_{\boldsymbol{i},\boldsymbol{j}}, e_\alpha)$ be given by the distance between the parallel supporting hyperplanes H_1 and H_2. The distance between H_1 and H_2 is equal to the distance between $H_1 \cap A$ and $H_2 \cap A$ where A is any linear subspace containing the normal vector of H_1 and H_2. Thus, let $A = \operatorname{span}\{v_{j_1}, \ldots, v_{j_\alpha}\} \ni e_\alpha$. $G_{\boldsymbol{i},\boldsymbol{j}}$ is contained in a parallelotope, $G_{\boldsymbol{i},\boldsymbol{j}} \subseteq \cap_{\beta=1}^{k} \tilde{H}_{j_\beta}$ where $\tilde{H}_{j_\beta} = \left\{x \in \mathbb{R}^d : \delta_{i_\beta} \le \left\langle x, v_{j_\beta} \right\rangle \le \delta_{i_\beta+1}\right\}$. Let $P = \cap_{\beta=1}^{\alpha} \tilde{H}_{j_\beta} \cap \operatorname{span}\{v_{j_1}, \ldots, v_{j_\alpha}\}$. Then P is a parallelotope contained in the α-dimensional vector space $V = \operatorname{span}\{v_{j_1}, \ldots, v_{j_\alpha}\}$ with widths $w(P, v_{j_\beta}) = \delta_{i_\beta+1} - \delta_{i_\beta}$, for $\beta = 1, \ldots, \alpha$. Thus we can apply Lemma 4.1 and conclude that

$$w(G_{\boldsymbol{i},\boldsymbol{j}}, e_\alpha) \le w(P, e_\alpha) \le \alpha!(\delta_{i_\alpha+1} - \delta_{i_\alpha}) \text{ for } \alpha = 1, \ldots, k.$$

For the first inequality, we use the fact that $w\left(\cap_{\beta=1}^{\alpha} \tilde{H}_{j_\beta}, e_\alpha\right) \ge w\left(\cap_{\beta=1}^{k} \tilde{H}_{j_\beta}, e_\alpha\right)$, and that $w\left(\cap_{\beta=1}^{\alpha} \tilde{H}_{j_\beta}, e_\alpha\right) = w(P, e_\alpha)$ since the distance between any two supporting hyperplanes H_1 and H_2 of $\cap_{\beta=1}^{k} \tilde{H}_{j_\beta}$ is equal to the distance between $H_1 \cap A$ and $H_2 \cap A$ where A is any linear subspace containing the normal vector of H_1 and H_2. ☐

We will rely on the following representation for a k-dimensional parallelotope. For sets A and B, let $A + B = \{a + b : a \in A, b \in B\}$.

Lemma 4.2. *Let V be a k-dimensional vector space, and $P := \cap_{\beta=1}^{k} \tilde{E}_\beta$ be a parallelotope where $\tilde{E}_\beta := \{x \in V : 0 \le \langle x, v_\beta \rangle \le d_\beta\}$ for k linearly independent normal unit vectors v_β. Let $\tilde{H}_\beta^+ := \{x \in V : \langle x, v_\beta \rangle = d_\beta\}$. Let $\tilde{f}_\beta$ be the unit vector lying*

in $\cap_{\gamma=1,\gamma\neq\beta}^{k}\tilde{H}_{\beta}^{+}$ *with* $\left\langle \tilde{f}_{\beta}, v_{\beta}\right\rangle > 0$*, for* $\beta = 1,\ldots,k$*. Then* 0 *is a vertex of* P *and we can write*

$$P = \sum_{\beta=1}^{k}[0, f_{\beta}]$$

where $f_{\beta} := d_{\beta}\tilde{f}_{\beta}/\left\langle \tilde{f}_{\beta}, v_{\beta}\right\rangle$*,* $[0, f_{\beta}] = \{\lambda f_{\beta} : \lambda \in [0,1]\}$*.*

Proof. Let $\tilde{H}_{\beta}^{-} := \{x \in V : \langle x, v_{\beta}\rangle = 0\}$. Since the vectors v_{β} are unique, $\cap_{\beta=1}^{k}\tilde{H}_{\beta}^{-} = 0$ and the intersection of any $k-1$ of the hyperplanes $\tilde{H}_{\beta}^{-}$ gives a 1-dimensional space, $\mathrm{span}\left\{\tilde{f}_{\beta}\right\}$. A k-dimensional parallelotope can be written as the set-sum of the k intervals emanating from the vertex, each given by the intersection of $k-1$ of the hyperplanes H_{β}^{-}. See page 56 of Grünbaum (1967). Note that f_{β} satisfy $\langle f_{\beta}, v_{\beta}\rangle = d_{\beta}$ so that $\tilde{f}_{\beta} \in H_{\beta}^{+}$; thus the k intervals are given by $[0, f_{\beta}]$, $\beta = 1,\ldots,k$. □

The next proposition combines the previous ones to bound the widths of $G_{\boldsymbol{i},\boldsymbol{j}}$ (i.e., to embed $G_{\boldsymbol{i},\boldsymbol{j}}$ in a hyperrectangle).

Proposition 4.2. *For each* $k \in \{1,\ldots,d-1\}$*,* $\boldsymbol{i} \in I_k, \boldsymbol{j} \in J_k$*, and each* $G_{\boldsymbol{i},\boldsymbol{j}}$*, and the basis* $\mathbf{e}$ *from Proposition 3.1, for* $\alpha = k+1,\ldots,d$*, we have*

$$w(G_{\boldsymbol{i},\boldsymbol{j}}, e_{\alpha}) \leq 2L_{k,1}w(G_{\boldsymbol{j}}, e_{\alpha}) \tag{19}$$

and

$$\mathrm{Vol}_d(G_{\boldsymbol{i},\boldsymbol{j}}) \leq (2L_{k,1})^{d-k}\,\mathrm{Vol}_{d-k}(G_{\boldsymbol{j}}) \cdot \prod_{\alpha=1}^{k}\frac{\delta_{i_{\alpha}+1} - \delta_{i_{\alpha}}}{\left\langle \tilde{f}_{\alpha}, v_{j_{\alpha}}\right\rangle} \tag{20}$$

where $L_{k,1}$ *is given by (22) and* $\tilde{f}_{\alpha}$ *is the unit vector with* $\left\langle \tilde{f}_{\alpha}, v_{j_{\alpha}}\right\rangle > 0$ *lying in* $\mathrm{span}\{v_{j_1},\ldots,v_{j_k}\}\cap\left(\cap_{\gamma=1,\gamma\neq\alpha}^{k}H_{\gamma}^{+}\right)$*,* $\alpha = 1,\ldots,k$*, where* $H_{\gamma}^{+} := \left\{y \in \mathbb{R}^d : \left\langle y, v_{j_{\gamma}}\right\rangle = \delta_{i_{\gamma}+1} - \delta_{i_{\gamma}}\right\}$*.*

Proof. Take $k \in \{1,\ldots,d-1\}$. Let x be an arbitrary fixed point, which we take to be $x \equiv x_{\boldsymbol{j}}$ (from (4)) for definiteness. Let $z = x + \sum_{\gamma=1}^{k} f_{j_{\gamma}}$ where $f_{j_{\gamma}} = d_{j_{\gamma}}\tilde{f}_{j_{\gamma}}$ where

$$0 \leq d_{j_{\gamma}} \leq (\delta_{i_{\gamma}+1} - \delta_{i_{\gamma}})/\left\langle \tilde{f}_{j_{\gamma}}, v_{j_{\gamma}}\right\rangle \tag{21}$$

and $\tilde{f}_{j_{\gamma}}$ is given by Lemma 4.2 for the k linearly independent normal vectors $v_{j_1},\ldots,v_{j_k}$. Take an arbitrary $e \in \mathrm{span}\{e_{k+1},\ldots,e_d\}$. Let $\lambda > 0$ be such that $\left\langle z + \lambda e, v_{j_{\beta}}\right\rangle$ is maximal over $\beta \in \{k+1,\ldots,N\}$ where $\lambda > 0$ is such that $z + \lambda e$ is in the boundary of $G_{\boldsymbol{i},\boldsymbol{j}}$. (That is, $v_{j_{\beta}}$ corresponds to the first hyperplane $z + \tilde{\lambda}e$ intersects for $\tilde{\lambda} > 0$.) Note that this means $\left\langle v_{j_{\beta}}, e\right\rangle < 0$. Then $\left\langle z + \lambda e, v_{j_{\beta}}\right\rangle = p_{j_{\beta}} + u$ so, with the δ sequence in (7) and $G_{\boldsymbol{i},\boldsymbol{j}}$ defined for any $u > 0$, we have

$$\lambda = \frac{p_{j_{\beta}} + u - \left\langle z, v_{j_{\beta}}\right\rangle}{\left\langle e, v_{j_{\beta}}\right\rangle} \leq \frac{\left\langle x, v_{j_{\beta}}\right\rangle - p_{j_{\beta}} + u\sum_{\gamma=1}^{k}\frac{\left\langle \tilde{f}_{\gamma}, v_{j_{\beta}}\right\rangle}{\left\langle \tilde{f}_{\gamma}, v_{j_{\gamma}}\right\rangle}}{\left\langle -e, v_{j_{\beta}}\right\rangle}.$$

Let

$$L_{k,1} := \sup_{e \in \operatorname{span}\{e_{k+1},\dots,e_d\}} 1/\left\langle -e, v_{j_\beta}\right\rangle, \tag{22}$$

which is finite since $G_{\boldsymbol{j}}$ is bounded. Then

$$\left\langle x, v_{j_\beta}\right\rangle - p_{j_\beta} \le d(x, H_{j_\beta}) \le d(x, \partial G_{\boldsymbol{j}}, e)$$

since H_{j_β} is the closest hyperplane to x in the direction e. Now, by (5) and (6), we have shown

$$\lambda \le 2L_{k,1} d^+(x, \partial G_{\boldsymbol{j}}, e), \tag{23}$$

meaning that

$$(G_{\boldsymbol{i},\boldsymbol{j}} - z) \cap \operatorname{span}\{e_{k+1},\dots,e_d\} \subset 2L_{k,1}\left(G_{\boldsymbol{j}} - x\right)$$

so we can conclude that $w(G_{\boldsymbol{i},\boldsymbol{j}} - z, e_\alpha) \le 2L_{k,1} w(G_{\boldsymbol{j}}, e_\alpha)$ and $w(G_{\boldsymbol{i},\boldsymbol{j}}, e_\alpha) \le 2L_{k,1} w(G_{\boldsymbol{j}}, e_\alpha)$ since $\langle z, e_\alpha\rangle = 0$ for all d_{j_γ} given by the range (21), $\alpha = k+1,\dots,d$, for $k = 1,\dots,d-1$. It then also follows that

$$\operatorname{Vol}_d(G_{\boldsymbol{i},\boldsymbol{j}}) \le (2L_{k,1})^{d-k} \operatorname{Vol}_{d-k}\left(G_{\boldsymbol{j}}\right) \cdot \operatorname{Vol}_k\left(\sum_{\alpha=1}^{k} [0, f_\alpha]\right), \tag{24}$$

where $f_\alpha = (\delta_{i_\alpha+1} - \delta_{i_\alpha})\tilde{f}_\alpha / \left\langle \tilde{f}_\alpha, v_{j_\alpha}\right\rangle$ and $\tilde{f}_\alpha$ given in the statement of the proposition. This yields (20). $\square$

Lemma 4.3. *Let Assumption 1 hold and let $G_{\boldsymbol{i},\boldsymbol{j}}$ be as in (8). If $G_{\boldsymbol{j}} = \emptyset$, then $G_{\boldsymbol{i},\boldsymbol{j}} = \emptyset$.*

Proof. This follows from Proposition 4.2 and its proof. $\square$

The above provides a hyperrectangle containing $G_{\boldsymbol{i},\boldsymbol{j}}$. Let $A+B = \{a+b : a \in A, b \in B\}$ for sets A, B. Let $\rho_{\boldsymbol{j},\alpha} := w(G_{\boldsymbol{j}}, e_\alpha)$ and then let

$$R_{\boldsymbol{i},\boldsymbol{j}} := \sum_{\alpha=1}^{k} [\alpha!(\delta_{i_\alpha+1} - \delta_{i_\alpha})e_\alpha, \alpha!(\delta_{i_\alpha+1} - \delta_{i_\alpha})e_\alpha] + \sum_{\alpha=k+1}^{d} [-2L_{k,1}\rho_{\boldsymbol{j},\alpha}e_\alpha, 2L_{k,1}\rho_{\boldsymbol{j},\alpha}e_\alpha]. \tag{25}$$

Then, for any $x \in G_{\boldsymbol{i},\boldsymbol{j}}$ we have shown

$$G_{\boldsymbol{i},\boldsymbol{j}} \subset x + R_{\boldsymbol{i},\boldsymbol{j}}. \tag{26}$$

References

Ball, K. (1992). Ellipsoids of maximal volume in convex bodies. *Geometriae Dedicata*, **41** 241–250.

Ball, K. (1997). An elementary introduction to modern convex geometry. In *Flavors of Geometry* (S. Levy, ed.), vol. 31. Flavors of geometry, 1–58.

Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probability theory and related fields*, **97** 113–150.

Bronshtein, E. M. (1976). *epsilon*-entropy of convex sets and functions. *Siberian Mathematical Journal*, **17** 393–398.

Dey, T. K. and Pach, J. (1998). Extremal problems for geometric hypergraphs. *Discrete & Computational Geometry*, **19** 473–484.

Doss, C. R. and Wellner, J. A. (2015a). Global rates of convergence of the mles of log-concave and s-concave densities. *arXiv:1306.1438*. 1306.1438.

Doss, C. R. and Wellner, J. A. (2015b). Inference for the mode of a log-concave density. Tech. rep., University of Washington. In preparation.

Dryanov, D. (2009). Kolmogorov entropy for classes of convex functions. *Constr. Approx.*, **30** 137–153.

Dudley, R. M. (1978). Central limit theorems for empirical measures. *The Annals of Probability*, **6** 899–929.

Dudley, R. M. (1984). A course on empirical processes. In *École d'été de probabilités de Saint-Flour, XII—1982*, vol. 1097 of *Lecture Notes in Math.* Springer, Berlin, 1–142.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Grünbaum, B. (1967). *Convex Polytopes*. With the cooperation of Victor Klee, M. A. Perles and G. C. Shephard. Pure and Applied Mathematics, Vol. 16, Interscience Publishers John Wiley & Sons, Inc., New York.

Guntuboyina, A. and Sen, B. (2013). Covering numbers for convex functions. *IEEE Transactions on Information Theory*, **59** 1957–1965.

Guntuboyina, A. and Sen, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probability theory and related fields,* to appear.

John, F. (1948). Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*. Interscience Publishers, Inc., New York, N. Y., 187–204.

Kim, A. K. H. and Samworth, R. J. (2014). Global rates of convergence in log-concave density estimation. *arXiv:1404.2298v1*. 1404.2298v1.

Koenker, R. and Mizera, I. (2010). Quasi-concave density estimation. *Ann. Statist.*, **38** 2998–3027.

Kolmogorov, A. N. and Tihomirov, V. M. (1961). ε-entropy and ε-capacity of sets in function spaces. *AMS Translations: Series 2*, **17** 277–364.

Rothschild, B. L. and Straus, E. G. (1985). On triangulations of the convex hull of n points. *Combinatorica*, **5** 167–179.

Seijo, E. and Sen, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *The Annals of Statistics*, **39** 1633–1657.

Seregin, A. and Wellner, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.*, **38** 3751–3781. With supplementary material available online.

van de Geer, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ Pr.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer-Verlag, New York.

Xu, M., Chen, M. and Lafferty, J. (2014). Faithful variable screening for high-dimensional convex regression. *arXiv.org*. 1411.1805v1.